DOCUMENT RESUME

ED 466 009 EA 031 661

AUTHOR Mitchell, Douglas E.; Mitchell, Ross E.
TITLE Evaluating the Impact of California's Class Size Reduction Initiative on Student Achievement: Third Year Evaluation Report.
INSTITUTION California Educational Research Cooperative, Riverside.
PUB DATE 2001-00-00
NOTE 28p.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Class Size; Elementary Education; Program Effectiveness; Program Evaluation; School Effectiveness; Teacher Student Ratio
IDENTIFIERS *California; *Class Size Reduction

ABSTRACT

This report reviews data generated during the first 3 years of California's Class Size Reduction (CSR) implementation. It focuses on a subpopulation of students who have had anywhere from zero to 3 years of reduced class-size experience, beginning in the first, second, or third grades, with some having returned to large classes in the fourth grade. The analysis is based on complete student, classroom, and teacher records from 15,267 third- and fourth-grade students in 546 classrooms from 72 schools in 7 southern California school districts. The data include reading, mathematics, and language test scores from the Stanford Achievement Test, as well as 36 variables covering student demographics, school assignments, classroom contexts, and teacher characteristics. The review found that since CSR was fully implemented the children first given exposure to smaller classes were not a representative sample of California's public school children. Positive achievements in mathematics accompanied CSR students, whereas language arts and reading showed trivial improvement. The report cautions that it is difficult to assess CSR's impact for a number of reasons, particularly the fact that CSR was initiated in concert with many other reforms, making it impossible to assess which changes resulted in student gains. Detailed information on the study sample is provided. (Contains 44 references, 7 tables, and 3 figures.) (RJM)

Evaluating the Impact of California's Class Size Reduction Initiative on Student Achievement: Third Year Evaluation Report

Douglas E. Mitchell
Ross E. Mitchell
California Educational Research Cooperative
University of California, Riverside

I. Introduction

This is the second in a series of California Educational Research Cooperative (CERC) reports analyzing how California's Class Size Reduction (CSR) initiative has impacted student achievement. This report reviews data generated during the first three years of CSR implementation. It focuses on a sub-population of students who have had anywhere from zero to three years of reduced class size experience, beginning in the first, second, or third grades, with some having returned to large classes in the fourth grade. The analysis is based on complete student, classroom and teacher records from 15,267 third and fourth grade students in 546 classrooms from 72 schools in 7 Southern California school districts. The data include reading, mathematics and language test scores from the Stanford Achievement Test (9th Edition – SAT-9) collected through California's STAR testing program. Also analyzed are 36 variables covering student demographics, school assignments, classroom contexts, and teacher characteristics. This introduction reviews the development of CSR in California and examines recent research undertaken by others supporting seven broad conclusions about how CSR programs are affecting student achievement.

Following this introduction is a summary of the key findings from this study (Section II). An overview of factors limiting our ability to isolate the effects of CSR is presented in Section III. Section IV describes the design of the CERC study and Section V documents the representativeness of the study sample. Section VI details the major findings from this research project, while Section VII concludes with a report on factors other than CSR which strongly influence student achievement.

A. Class Size Reduction is a very expensive policy, making careful evaluation of its potential benefits very important.

In California, the *Class Size Reduction Program* authorized by Senate Bill 1777 in 1996 continues to represent the most expensive educational reform effort ever undertaken by any state. State funds allocated during the first three years of operation amounted to nearly $4.1 billion – about $3.3 billion for operation, with an additional $0.8 billion required for school facilities (Table 1). These figures do not include any expenditures from local school district general funds that may have been needed to offset excess staff or facilities costs (for other state and national figures and estimates see Brewer, Krop, Gill, & Reichardt, 1999; Hertling, Leonard, Lumsden, & Smith, 2000; National Conference of State Legislatures, 1998).

Table 1. State Funding Allocations by Category for the *California Class Size Reduction Program*, grades K-3, from 1996-1997 through 1998-1999

School Year	Operations	Facilities	Total	Cumulative Total
1996-1997	$611,275,000	$342,802,500	$954,077,500	$954,077,500
1997-1998	$1,216,587,000	$311,628,438	$1,528,215,438	$2,482,292,938
1998-1999	$1,439,456,096	$154,360,000	$1,593,816,096	$4,076,109,034

Sources: The following documents were downloaded on January 22, 2001 from the California Department of Education Class Size Reduction website:
http://www.cde.ca.gov/classsize/particip/sum96.htm,
http://www.cde.ca.gov/classsize/particip/sum97.htm,
http://www.cde.ca.gov/classsize/particip/sum98.htm, and
http://www.cde.ca.gov/classsize/facts.htm

B. National interest in CSR remains high; new research has been published and important policy conferences have focused political attention.

While this report documents ongoing research work sponsored by the California Educational Research Cooperative (CERC), we note that other recent studies also offer important insights into the overall impact of class size on teacher behavior and student achievement. Tennessee *Project STAR* data continue to be reanalyzed, including various efforts to follow the 1985 cohort through later elementary, middle, and high school years (e.g., Blatchford, Goldstein, & Mortimore, 1998; Finn & Achilles, 1999; Finn, Gerber, Achilles, & Boyd-Zaharias, in press; Goldstein & Blatchford, 1997; Hanushek, 1999; Krueger, 1999; Krueger & Whitmore, 2001; Nye, Hedges, & Konstantopoulos, 1999, 2000; Pate-Bain, et al., 1997). These improvements and refinements reconfirm earlier analyses indicating that Tennessee's CSR was successful at facilitating improved student performance. But there is some reason to believe that the effects of CSR in Tennessee were slightly less powerful than originally reported. Additionally, the benefits of a small class experience for students who were not enrolled in the program until second or third grade are noticeably less than that obtained by those who started in kindergarten or first grade. Unfortunately, the single cohort design does note permit a clear distinction between the effects of student mobility and timing of CSR experience because too few students were permitted to violate the design by moving from a large class to a small class while remaining in the same school. Thus, with exhaustive reanalyses, the basic conclusions offered from *Project STAR* remain the same:

- earlier is better (K or first grade),
- longer is better (K/1 through third – at least three years – offers the greatest benefit),
- a more conducive classroom learning environment is produced, and
- positive student achievement, behavior and attitude effects persist, but weaken as students continue through school.

Other recent efforts worthy of attention include the Wisconsin *SAGE* study and a study in England examining class size and the adult-pupil ratio. The Wisconsin program has substantially

reproduced the basic outlines of the Tennessee studies: improvement in the first year with the improved performance remaining stable in subsequent years for students enrolled in a class with a reduced student to certified teacher ratio of 15 to 1 – this includes classes with two teachers and 30 students – and a greater benefit to African American students (Molnar, et al., 2000). These results are most notable for mathematics achievement, while benefits in reading and language are smaller. In an examination of the first three years of reporting on *SAGE,* Hruz (2000) cautions that the positive results may be due almost entirely to the benefit to African American students, however, since white students are not benefiting greatly if at all.

The Wisconsin evaluators are making some effort to attend to teacher disposition and work performance, but their study design does not permit them to make causal inferences about the link between teacher attitudes and behaviors and student outcomes as a function of class size. A point related to teaching that has not received much attention is that the classes identified as high performing have much higher average teacher experience than the low performing classes (Hruz, 2000). Thus, the question of whether it is the benefit of having experienced teachers or a reduced size class that is more strongly related to student achievement remains open.

The British study also confirms that small classes at the start of school are beneficial to students, and that initially low achieving students benefit most from the experience (Blatchford, 2000). Further, teacher ability and effort to attend to individual pupil needs and performance is increased in a reduced size class, where student attention is better maintained, and disruptive and off-task behavior is reduced. But an important cautionary note is offered in this study as well. The smaller class size creates a social environment that can lead to more aggressive children or to children being rejected by their peers. Either due to lack of alternative peers, or lack of a perceived need to interact with and learn from peers, the young English children in this study displayed more social adjustment difficulties. Thus, the story is fairly consistent outside of Tennessee, both within and outside of the United States. A small class experience is most effective when students begin school (K/1), most valuable to students who are at-risk, and the benefits are more likely to persist if students are in smaller classes longer. But despite the average gains associated with class size reduction, not all small classes are beneficial nor are all large classes detrimental.

C. In sum, work to date supports seven broad conclusions:

1. The overall effects of CSR are modest in size, and in danger of being obscured by other factors influencing student achievement

2. Earlier exposure to CSR is more likely to produce significant achievement gains.

3. Longer participation in small classes does not necessarily produce greater achievement gains, but may make the gains more resistant to decay.

4. The effects of small class experience persist after children return to larger classes, but these effects tend to decay over time.

5. Some populations of students seem to gain more from participation in small classes than others. Specifically, at-risk poor and under-represented minority children seem to show slightly larger gains for the same amount of exposure.

6. While classroom processes and curriculum content are certainly important factors in achievement, it is hard to document specific changes in instruction that are related with reductions in class size.

7. The 1999 finding by CERC researchers that California's Class Size Reduction Initiative produced vanishingly small impact on student achievement as measured by the mandated Stanford Achievement Test – 9th Edition was confirmed by a substantially funded statewide CSR evaluation consortium (Bohrnstedt & Stecher, 1999).

II. Summary of Findings in this Report

This report documents the five important new findings regarding the first three years of experience under California's Class Size Reduction Initiative. These findings include:

A. Since California's CSR initiative was implemented as a fully operational program, rather than an experimental or test program, the children first given exposure to the small classes were ***not*** a representative sample of California's public school children. Data show substantial differences in such factors as: the number of overage children in the classroom, the use of multi-track year-round calendars, the amount of experience of the teachers, the incidence of special education students in the classrooms and the proportion of African American students in the classes. There are smaller, but potentially important, differences in other factors that are strongly associated with student achievement like the socio-economic status of children and their home languages (additional documentation of implementation difference can be found in Bohrnstedt & Stecher, 1999; D. Mitchell & R. Mitchell, 1999; Stecher & Bohrnstedt, 2000).

B. The picture for mathematics is quite different this year from that found in our previous report. The data indicate positive achievement gains in this subject area, but still document only trivial impacts in reading and language arts. There is an important cautionary note on this finding, however. All of our third grade students had at least some small class experience and the large gains in mathematics were among those third grade students. Thus, the math achievement gains may reflect age-cohort differences rather than CSR impact.

C. The tendency for benefits to accrue primarily to children who have their reduced size class experience during their earliest school years is documented in detail. Indeed, children starting their CSR experience in the third grade actually show some loss in academic attainment.

D. We found continued support for the conclusion that the positive benefits accruing to students through participation in small classes are weaker than the negative consequences of at least four factors that tend to interfere with academic learning. These factors include:

1. Poverty
2. Coming from a family that does not speak English at home
3. Being of African American or Hispanic ethnic heritage
4. Being in a special education Resource Specialist Program.

E. Data collected for this project show that at least 40 percent of the variability in student achievement (at least the kind measured by the SAT-9) is governed by factors that have nothing to do with smaller classes, variations in curriculum, teachers' instructional practices, school or district curriculum policies, or with children's ability, engagement in school and prior academic attainment. Student demographics, classroom organization and teacher training and experience account for this 40 percent of student achievement differences.

III. A Cautionary Note: Accurate Assessment of CSR Impact is Quite Challenging

Five Problems are encountered whenever we try to evaluate broad policies like CSR. They include:

First, CSR is accompanied by a host of other efforts to improve achievement – the impacts of many of these efforts cannot be easily separated from the impact of changing class size. California enacted more than a dozen school reform and improvement policies during the same period as the development and implementation of CSR, including:

1) Passage of California Proposition 227 which has sharply curtailed bilingual education programs,
2) Adoption of a statewide accountability policy forcing multiple assessments of student achievement and requiring reports on all students not reaching grade-level achievement standards,
3) Implementation of a Beginning Teacher Support and Assessment program creating a two year induction program for new teachers,
4) Changes in the funding model for special education which substantially affects local district costs when children are certified for services,
5) Changing economic conditions that affect unemployment and poverty rates in many districts,
6) Continued immigration and relocation which changes the composition of many school populations,
7) A broad reading initiative aimed at changing the focus and effectiveness of early literacy instruction,

8) Changes in regulations regarding the certification of teachers that have changed both the character and timing of pre-service teacher preparation,
9) Support for development of new instructional technologies aimed at providing students with better access to location-independent and multi-media learning opportunities,
10) Adoption of a new statewide standardized achievement test (the Stanford Achievement Test, version 9) and mandated school level public reporting of achievement test scores,
11) Continued implementation of new textbook and curriculum materials adoption cycles (both language arts and mathematics curriculum frameworks were changed at the time of CSR policy adoption and implementation) assuring major changes in the scope, sequence and content of subject matter curricula,
12) Addition of ninth grade class size reduction for specific subjects,
13) Changes in regulations regarding the certification of school administrators that have changed both the character and timing of pre-service administrator preparation.
14) Establishment of a powerful Peer Assisted Review (PAR) program aimed at holding experienced teachers accountable for self-improvement.

Second, the impact of reducing class size is entangled with and embedded in a wide range of student demographic, classroom, school and district factors that have powerful effects on achievement making it impossible to make simple direct measurements of the specific contributions of CSR. As a result, statistical analysis has to be used to disentangle the several contributions to student achievement – but even the best statistical techniques do not give foolproof tests.

Among the most prominent demographic factors that are known to have effects large enough to obscure class size effects are: family poverty, ethnicity, home language, inter-school transiency and student gender (e.g., Entwisle & Alexander, 1992; Han & Hoover, 1994; Jencks & Phillips, 1998; D. Mitchell & R. Mitchell, 1999; Rosenthal, Baker & Ginsburg, 1983). Within schools, such factors as grade to grade cohort achievement variations, special education placements, language proficiency levels, combination grade class assignments, and grade-level retention can be expected to influence measured achievement (e.g., Balow & Schwager, 1990; Burns, 1996; Entwisle, Alexander, & Olson, 1997; Hakuta, Butler, & Witt, 2000; Mitchell, Karam, & Destino, 1998).

At the classroom level, achievement is influenced by such factors as: the use of year-round or traditional calendars, the willingness of schools to utilize combination grade classes to manage enrollments, and the extent to which students are segregated by socio-economic status, ethnicity, language fluency levels, ability, gender or special education category (e.g., Burns & Mason 1998; R. Mitchell & D. Mitchell 1999; Rowan & Miracle, 1983; Shields & Oberg, 2000; Veenman 1995). Any of these factors might obscure the effects of CSR.

Teacher assignments also vary from class to class. Confounded with class size reduction we are likely to find variations in teacher credentials, experience, age, contract status, ethnicity, gender

and educational attainment (e.g., Alexander, Entwisle, & Thompson 1987; Darling-Hammond, 1998; Wright, Horn, & Sanders 1997). Finally, school and district boundaries serve to segregate students by neighborhood, culture, socio-economic background and other factors that are not easily measured (e.g., Arum, 2000; Black, 1999; Clotfelter, 1998; Entwisle, Alexander, & Olson 1997). All of these factors need to be considered as possible sources of achievement variation before we can confidently conclude that students have benefited significantly from taking instruction in reduced size classes.

Third, while most attention is focused on the ***average*** level of achievement for all students experiencing smaller classes, it is not clear that this is the only or even the most important outcome of interest. CSR might be judged successful if it provided the benefits only to the children in greatest need of academic help; or it might be seen as a failure if it interfered with the achievement of specific groups.

If, for example, classroom averages remain relatively constant, but previously failing students are now meeting grade-level standards, would that suffice to justify the expense of this policy? Or, if class averages go up, but low attaining students are no better off than they were before, would that be considered a failure? If class averages go up, but the attainment of students is concentrated on the middle range, so that previously high attaining students are no longer moving ahead as rapidly, would that be considered a failure? In short, what ***patterns*** of classroom attainment are being generated, and how are those patterns to be evaluated?

Fourth, particularly in California, implementation procedures may have distorted the normal, long-term impact of CSR because schools had to find classroom space and new teachers on short notice in circumstances when both were in short supply. By the same token, if we put off assessing its impact until all implementation wrinkles are straightened out, it will be impossible to separate CSR from other factors affecting overall student achievement.

Since local school districts had to implement the policy in a matter of a few months, it was difficult to make needed changes in classroom space and teacher recruitment. Schools of education had no advanced warning, with the result that they prepared no surplus of new teachers to take up the large number of new teaching positions created. Construction companies did not have an opportunity to gear up for the production of new classroom facilities. Even if they did anticipate construction needs, there was no early release of construction funds to prepare classrooms. New teachers, not fully qualified teachers, and teachers transferring to new assignments at the last moment had to start instruction of smaller classes in new spaces. Sometimes such irregular spaces as libraries, multipurpose rooms or computer laboratories were converted for the new classes. A significant number of these problems continue into the second and subsequent years of implementation (Bohrnstedt & Stecher, 1999; Hymon, 1997; Illig, 1997; Ogawa & Stine, 1998; Stecher & Bohrnstedt, 2000; Wexler, et al., 1998).

Fifth, since California does not begin systematic achievement testing until the end of second grade, it is not possible to ascertain whether CSR in this state is having substantial impact during this first critical year of schooling.

Results from Tennessee's Project STAR indicate that the major effects of class size reduction are experienced during the kindergarten year, or during the first year a child is exposed to this form of instruction (e.g., Finn & Achilles 1990; Finn, Gerber, Achilles, & Boyd-Zaharias, in press; Krueger 1999; Nye, Hedges, & Konstantopoulos, 2000). If this is generally true, it may not be possible to measure the effects of class size reduction in settings like California where the small class experiences could begin in the first, second or third grade and may not be encountered by some children until their second or third year of schooling. Additionally, it is possible, that achievement gains produced during an initial exposure to small classes will not be sustained over time. Careful attention to this issue is required before the job of evaluation can be considered complete.

IV. The CERC Evaluation Study

This report assesses the educational experiences of third and fourth grade students in seven Southern California school districts. The district enrollments range in size from about 600 to nearly 36,000 and represent a broad cross-section of urban, suburban and rural settings. The study design has five important features:

A. The CERC study is longitudinal in nature, examining the ultimate achievement levels of students whose individual class size histories are known.

The analyses presented here are based on carefully tracing the experiences of students in school districts where, due to implementation decisions made by district leaders, both large and small classes were created for children in all of the target grades (kindergarten through grade three). All available records from students in regular classrooms (i.e., not community schools, individual tutorial students, special education Special Day Class classrooms, or combination grade classrooms with more than two grades) in each of the two study grades within each of the participating districts are included in this study. They consist of 15,267 third and fourth graders in 546 classrooms in 72 schools. The student records selected for analysis are those for which a three-year history of class size reduction experience could be determined, where complete matching of students with teachers could be made, and where complete data on student classroom assignments were available. Of the original sample, 2,964 students were lost due to incomplete data. The largest portion of the sample reduction is due to incomplete class size reduction experience histories, which is almost entirely due to inter-district mobility resulting in discontinuities in student records. (We were able to secure records for most students who moved between schools within the same district, which is the type of transient student retained in the analysis). Additional losses occurred due to incomplete student testing – some students did not take, complete, or provide valid responses for subject area sub-tests. The final sample for analysis consisted 11,716 students with complete data and a total reading subject score, 12,039 with complete data and a total mathematics subject score, and 11,943 with complete data and a total language subject score.

B. The data analysis examines six groups of third and fourth grade students who have had zero, one, two or three years of experience in small classes.

As indicated in Table 2, the CERC study sample contains students with six different patterns of exposure to reduced size classes. Among students currently in third grade, most (4,925) have had three years of CSR starting in first grade, a substantial group (937) had two years of exposure starting in grade 1 (but returned to large classes in third grade), and a moderate size group (469) had two years of exposure starting in grade 2.

Table 2. Class Size Reduction Experiences for Seven District Sample of 3rd and 4th Grade Students through the 1998-1999 School Year

		Current Grade in School		Total
		3	4	
CSR Experience	2 Years Starting 1st Grade	937		937
	3 Years Starting 1st Grade	4925		4925
	1 Year Starting 2nd Grade		1247	1247
	2 Year Starting 2nd Grade	469	801	1270
	1 Year Starting 3rd Grade		1783	1783
	None		2141	2141

Among fourth graders in the study sample, the largest group (2,141) had no CSR experience; none of the fourth graders had started their CSR exposure in grade 1. Substantial groups of fourth graders have had either one or two years of exposure starting in either second or third grade.

C. As the data analyses presented in this study will confirm the timing of exposure to CSR is quite important. Students whose earliest CSR experience was in the first grade showed quite different results from those whose initial exposure was in second or third grades (no students with their initial exposure in kindergarten were available for this study).

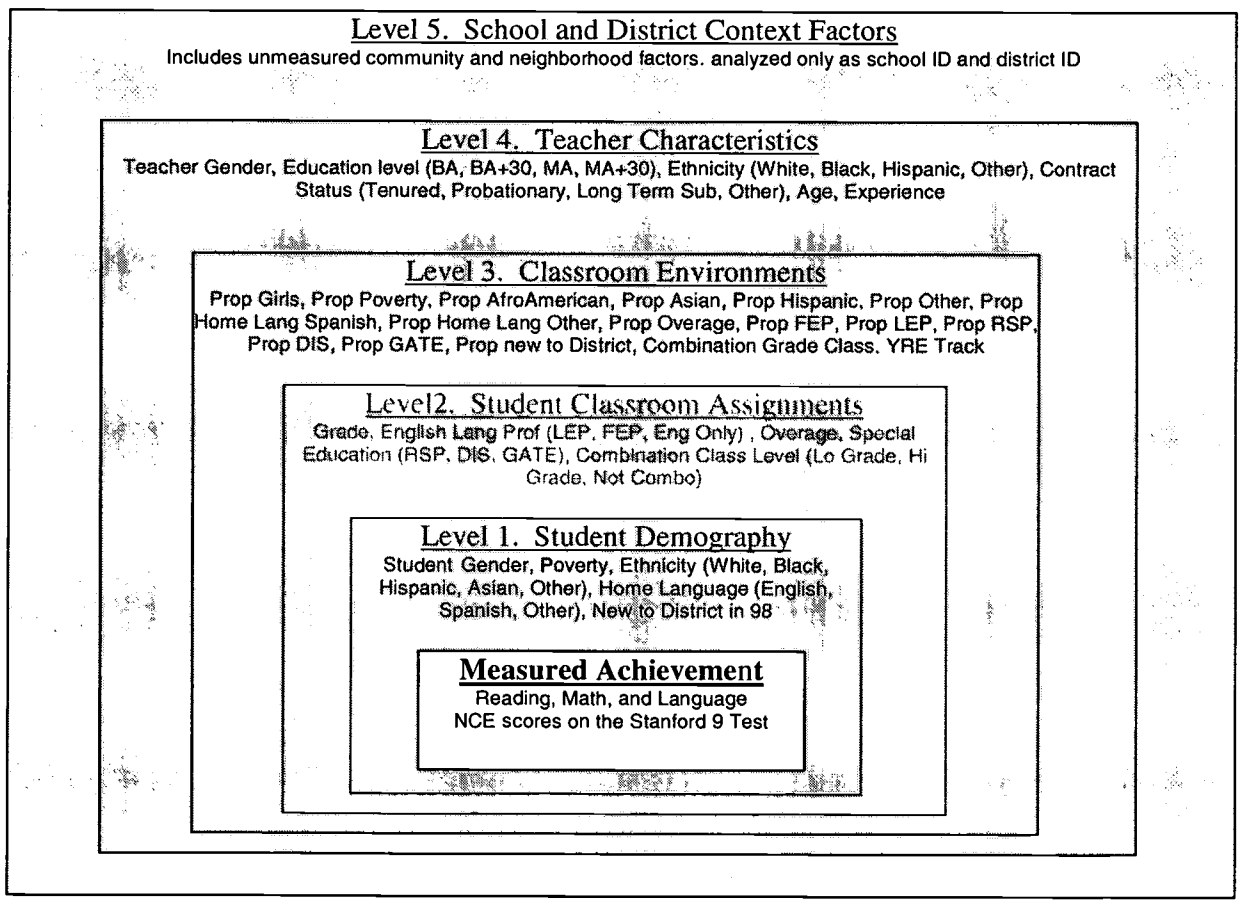
Exactly how important these differences are is hard to estimate because, by the third year of implementation, there were no students in the third grade who had spent no time in reduced size classes, and there were no students in the fourth grade who had started their CSR exposure in first grade.

D. Perhaps the most important feature of the CERC study design is that it allows us to estimate the confounding effects of a broad range of variables that could be masking the true effects of CSR experience.

Since California's CSR initiative was implemented as a rapidly expanding, full-fledged operational program, practical considerations made it inevitable that the children placed in small classes would not have the same demographic profiles, classroom contexts or prior achievement histories as those who continued to attend school in larger classes. By monitoring their complete demographic profiles, their teacher characteristics and their school program assignments, the CERC study is able to statistically control for these biasing factors and thus produce a much more accurate estimate of the true CSR impact on achievement.

E. Analysis procedures recognize that key variables operate at four distinct levels. There are: 1) individual and family factors, 2) classroom assignment variables, 3) teacher experience and demographic factors, and 4) school level variables. At each level, these variables may be exerting decisive influence on the capacity of CSR to materially influence achievement. A complete model of the variables studied is presented in Figure I.

The specific variables are described in detail in Appendix A.



V. The Study Sample

A. The students in the CERC study sample are quite representative of California's total public school student population.

Table 2a presents a statistical comparison of the 12,303 students and their classroom teachers in our sample with the 947,597 California students in grades 3 and 4. As shown at the top of the table, the two groups are very closely aligned on overall achievement in reading math and language. While the sample is generally representative of California's total school population, there are half a dozen places where the sample deviates substantially from the overall statewide population. As would be expected from the design of the study, which takes advantage of incomplete CSR implementation, the proportion of students in reduced size classes is somewhat below the state as a whole. Our study sample has more English home language students than the overall state population, with commensurately fewer Spanish and Other home language students. Correspondingly, there are fewer English Learner (LEP) students. Despite the high number of students in the study sample from low-income homes (NSLP eligible), the California proportion statewide is yet higher. White and Hispanic student populations are fairly similar, but the sample has a higher proportion of African/Black students and a lower proportion of Asian students. The sample also has less than half of the state's proportion of its students attending traditional calendar schools. This reflects an increasing use of the multi-track year-round calendar to create classroom space for CSR among the sample districts.

Table 2a also presents some descriptive statistics for the study sample on variables for which statewide population parameters were not available at the time this report was prepared. About 13 percent of the sample students are in combination grade classes. Nearly one out of every eleven students was new to the school where they were tested in 1999. Among year-round education tracks, Track C and Track D are the preferred ones. Together they enroll 21 percent more students than Tracks A and B. In our sample, there are only two year-round schools on 3-track attendance calendars, but the dates align perfectly in one case and nearly perfectly in another with three of the four tracks on the 4-track attendance calendars. As such, it is the attendance calendar for the track that determines each student's and teacher's designation.

Table 2b compares teacher characteristics in the CERC study sample with statewide averages. Teacher ethnic distribution reflects the student distribution reported in Table 2a. There are noticeably more African/Black teachers in the sample and fewer Other ethnicity teachers. The sample has an appreciably higher percentage of male elementary teachers for students in grades three and four than the overall state percentage for schools enrolling students in grade three and four. The proportion of fully credentialed teachers is only slightly higher than that of the state as a whole. The sample also has less than 15 percent more probationary teachers than the state population, matched by a reduction in the number of teachers having tenure contracts. Though the distribution differs, the total number of teacher on "temporary" or "other" contracts is nearly the same. There are a few percent more teachers with 30 semester hours beyond the bachelor's degree, matched by a reduction in those holding just the bachelor's degree. Finally, the average

teacher experience level for the grades three and four classrooms in this sample is more than a year-and-a-half lower than that for the state.

Table 2a. Comparison of Mean Achievement and Percentage of Students by Level for Each Factor for Seven District Sample with Elementary Schools Enrolling Third and Fourth Grade Students in the State of California in 1998-1999.

Factor	Levels	Sample	State
Mean SAT-9 Subject Total Achievement (NCE)	Reading	44.2	45.2
	Mathematics	48.3	48.3
	Language	47.2	47.2
Grade	3	51.3	51.5
	4	48.7	48.5
CSR Option 1 in 1996-97 (Grades 1-2)	Yes	63.9	71.5
	No	36.1	28.5
CSR Option 1 in 1997-98 (Grades 2-3)	Yes	72.5	80.7
	No	27.5	19.3
CSR Option 1 in 1998-99 (Grades 3-4)	Yes	44.2	44.0
	No	55.8	56.0
Student Ethnicity	African/Black	14.2	9.0
	Asian	3.2	7.5
	Hispanic	45.2	42.7
	White	35.1	36.8
	Other	2.3	3.9
Student Home Language	Spanish	23.5	31.0
	English	73.3	60.1
	Other	3.2	8.9
Student Low Income Status (NSLP Qualified)	Yes	47.1	56.0
	No	52.9	44.1
Student Gender	Male	50.6	51.0
	Female	49.4	49.0
Student Intra-District Mobility 1998-1999	New to School	9.3	N/A
	Not New	90.7	
Student English Language Proficiency	LEP	16.7	30.2
	FEP	10.0	9.7
	English Only	73.3	60.1
Student Special Education/ GATE	RSP	3.8	N/A
	DIS	2.1	
	GATE	9.8	
	Not Spec Educ	84.2	
Student Overage for Grade (15+ Months)	Overage	2.4	N/A
	Not Overage	97.6	
Student Grade in Combination Grade Classroom	Low Grade	7.0	N/A
	High Grade	5.7	
	Single Grade	87.3	
School Attendance Calendar	Traditional	41.9	86.6
	YRE A-Track	12.9	12.4
	YRE B-Track	13.3	
	YRE C-Track	16.1	
	YRE D-Track	15.7	
	YRE Single Track	0.0	1.1

Table 2b. Comparison of Percentage of Teachers by Level for Each Factor and Average Teaching Experience for Seven District Sample with Elementary Schools Enrolling Third and Fourth Grade Students in the State of California in 1998-1999.

Factor	Levels	Sample	State
Teacher Ethnicity	African/Black	7.6	4.9
	Hispanic	12.1	14.3
	White	77.3	74.1
	Other	3.0	6.7
Teacher Gender	Male	20.8	14.7
	Female	79.2	85.3
Teacher Credential Status	Full Credential	88.1	86.6
	Not Full Credential	11.9	13.4
Teacher Contract Status	Long-Term Sub/Temp	2.0	8.0
	Probationary	23.0	20.4
	Tenure	62.3	64.7
	Other	12.7	7.0
Teacher Education Level	MA & Up	25.6	26.1
	BA + 30	55.2	51.3
	BA	19.2	22.7
Avg. Teacher Experience (Years)		10.2	11.9

C. The final dataset was produced by combining SAT-9 data with CBEDS-PAIF data and retaining for study all students in grades three and four for whom it was possible to document their entire history of CSR participation.

These data, plus information on district CSR implementation generated the fifteen control variables described in Appendix A. Calculating classroom and school averages for these variables created an additional 16 context and control variables.

D. As described in the design section above, the final sample consists of six groups of students with differing combinations of starting grade and duration of exposure to CSR (see Table 2).

Two tiny groups of very exceptional students were dropped from the study because they consisted of retained students or those taking fourth grade instruction in a small 3-4-combination grade class.

VI. The Major Findings

As briefly summarized in Section II above, the CERC study of the third year of operation of California's CSR initiative has reached five basic conclusions. They are:

Conclusion #1: CSR implementation provided different groups of students with very different types of exposure to smaller classes, making statistical control over a wide variety of confounding variables absolutely essential if we are to discover the true effects of small class exposure on achievement.

Table 3 summarizes the some of the most obviously confounded variables that could easily obscure the impact of CSR on student achievement. Each of the factors in this list are significantly related to achievement and all interact with each other in complex and sometimes unpredictable ways.

Table 3

Factors Related to Achievement that Confound Analysis of CSR Impacts

School proportions of Confounding Factors	Differing CSR Experiences						Average all groups	Size of F-Statistic	Range Low	Range High
	2 Yrs - 1st	3 Yrs - 1st	1 Yr - 2nd	2 Yrs - 2nd	1 Yr - 3rd	No CSR				
Overage 15 or more months	4.2%	2.6%	2.2%	3.7%	4.1%	1.9%	2.9%	327.02	1.9%	4.2%
Special Ed: DIS	2.8%	1.9%	1.9%	0.9%	1.3%	2.6%	1.9%	241.00	0.9%	2.8%
Special Ed: RSP	4.9%	4.3%	4.3%	5.0%	5.0%	3.9%	4.4%	197.57	3.9%	5.0%
Year Round	19.2%	63.5%	65.0%	69.7%	54.1%	50.1%	57.2%	173.89	19.2%	69.7%
Teacher Experience	8.3	10.7	10.2	11.4	11.6	10.1	10.6	171.00	8.3	11.6
Teachers with tenure	57.7%	63.9%	61.5%	69.1%	71.9%	60.6%	64.3%	165.43	57.7%	71.9%
White ethnicity	25.1%	35.5%	29.5%	34.1%	43.0%	36.8%	35.2%	145.57	25.1%	43.0%
Teachers with full credential	82.6%	88.2%	88.7%	91.6%	91.3%	87.2%	88.4%	120.22	82.6%	91.6%
Students new to the school	22.1%	23.8%	27.0%	24.7%	18.9%	21.0%	22.9%	100.01	18.9%	27.0%
English Home Language	71.0%	73.4%	71.2%	76.1%	80.3%	73.4%	74.2%	95.28	71.0%	80.3%
School size	759	739	777	709	693	763	739	87.39	693	777
Teachers with less than BA+30	23.8%	18.8%	18.4%	20.5%	18.3%	17.8%	19.1%	36.97	17.8%	23.8%
Special Ed: GATE	6.6%	7.9%	7.7%	8.4%	8.4%	9.3%	8.2%	34.42	6.6%	9.3%
Students in Combo Classes	11.3%	15.5%	16.9%	14.8%	14.0%	15.8%	15.1%	30.79	11.3%	16.9%
Poverty	45.8%	45.6%	46.7%	40.1%	42.5%	46.7%	44.9%	25.12	40.1%	46.7%
Teachers with MA or above	24.1%	25.7%	27.0%	27.6%	25.2%	24.7%	25.7%	21.76	24.1%	27.6%

Each of the variables in Table 3 represents a school-wide average, measuring the school context within which CSR implementation has taken place rather than the characteristics of individual students. Though all are show statistically reliable differences across the six different class size experience groups, some are much more deeply entangled in CSR implementation than others. The sixteen variables in this table are ordered from the most powerful (the percent of children in the school who are 15 or more months overage for their grade) to the least powerful (the percent of teachers with holding advanced degrees) predictor of CSR experience.

Some of the variables most strongly associated with CSR implementation (like Year Round calendars and teacher tenure) are not as powerfully linked to achievement as some variables near the bottom of Table 3 which exert a lot of influence on achievement. Poverty, for example, is only modestly linked to CSR implementation, but it has a powerful effect on achievement. Home language and participation in GATE programs are also strongly associated with achievement but only moderately associated with CSR implementation. But even moderate confounding with CSR can mean substantial differences between groups of children with different CSR exposure. Poverty rates, for example, range from a low of 40.1% of the students in schools where children had two years of CSR starting in the second grade to a high of 46.7% of those who had no CSR exposure or those who had only one year starting in the second grade.

Average teacher experience varies by more than 3 years, and the percent of children in Year Round schools varied dramatically – the lowest rate was only 19.2 percent for children getting two years of CSR starting in the first grade to a high of 69.7 percent of those who got two years

of CSR starting in the second grade. And the non-English speaking proportion of the student population varied from less than 20 percent (80.3% English home language) to nearly 30 percent (71.0% English home language) for students with two years of CSR starting in grade one.

Examples could be multiplied endlessly here. The basic point is that implementation of CSR is deeply entangled with other variables known to influence student achievement. Taken together, these confounding variables are about 47 percent accurate in identifying the type of CSR experience children have had – making it abundantly clear that they are at least as likely to be the causes of achievement variations among CSR implementation groups as are the small classes themselves.

The only sensible way to proceed is to remove the effects of these confounding variables ***before*** trying to assess the impact of CSR. This is done using a statistical regression procedure that, in effect, equalizes the different CSR treatment groups on these variables before testing to see whether the groups, so conditioned, have significantly different achievement test scores.

> Conclusion #2: After controlling for critical confounding variables, mathematics achievement test scores show the only substantial largest benefit from CSR experience.

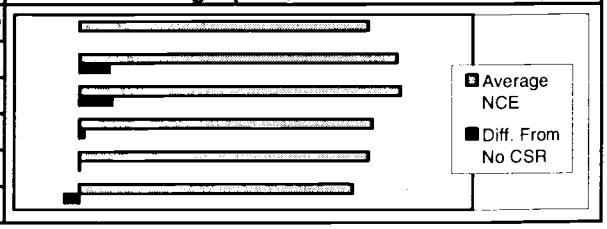
Because mathematics achievement scores show a different, and more promising pattern this year, this report will concentrate on analyzing that subject. Impacts on reading and language achievement are vanishingly small and will be discussed primarily to highlight the significance of the mathematics findings.

Jumping right to the infamous "bottom line," Table 4 reports the relative mathematics achievement gains for students with various combinations of CSR when compared to the 2,141 students in the study sample who had no CSR experience at all. The estimate achievement levels reported in the table are those that would be expected if all demographic, classroom assignment, classroom environment, teacher characteristics and school and district effects are statistically equalized for all students in the study group.

Table 4. Average SAT-9 Mathematics Achievement by Class Size Reduction Experience

(NCE scores adjusted for all known implementation biases)

Starting Grade	Number of Years	Average Math Score	Difference from No CSR	Bargraph of Test Scores
No CSR	Zero	43.95	0	
First	Two	48.42	4.47	
First	Three	48.94	4.99	
Second	One	44.62	0.67	
Second	Two	43.99	0.04	
Third	One	41.64	-2.31	



The table reports the average SAT-9 mathematics achievement for each CSR exposure group. Achievement is reported in Normal Curve Equivalent (NCE) scores, which have a nationally normed mean of 50, standard deviation of about 20. (A change of about 10 points represents one year of academic progress – this number varies from one grade to another). The actual mean for our sample was 48.3, a bit below the national mean but right in line with the California state mean. The standard deviation for our sample was 20.96, quite close to the expected value.

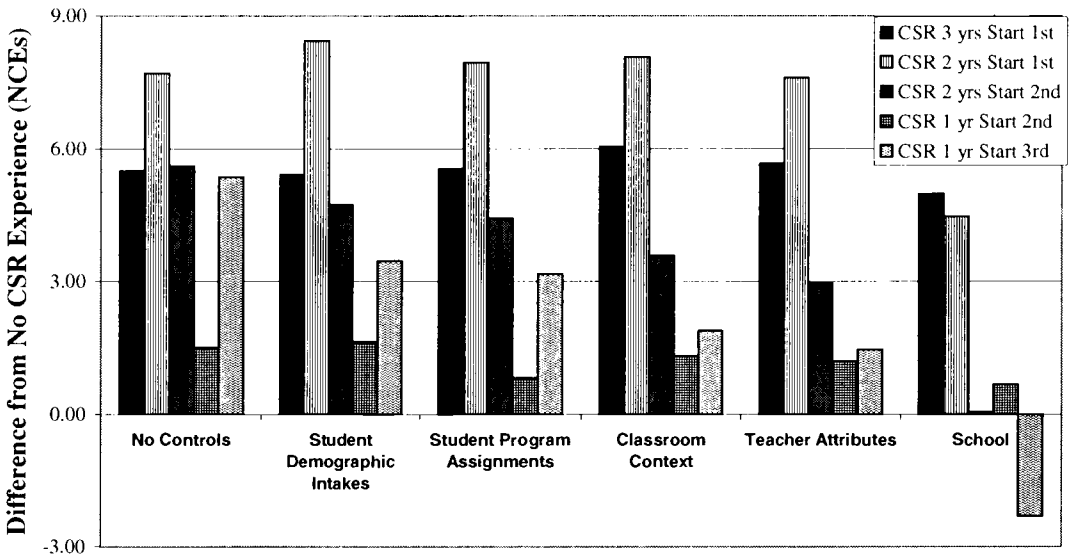
As shown at the top of the first column of numbers in the table, the estimated average for students who had no CSR exposure was 43.95, well below the 48.3 overall average. The other numbers in this column of the table report the average achievement for the groups of students with each of the five different CSR experiences. Only those students who had their initial exposure to small classes during their first grade in school show any significant improvement in their mathematics achievement. Those who started CSR in the second grade scored virtually the same as students with no CSR experience, and those whose first exposure was in grade three actually did less well than if they had received no CSR exposure at all. Indeed, the average of all the CSR exposed student groups would be only 1.57 NCE points above the students with no CSR experience – about 5 weeks of normal academic progress.

The four and a half to five point advantage attained by the first grade exposure groups represents nearly a half year of academic advantage over their no CSR peers. This advantage, if it can be believed, compares favorably with that documented in Tennessee's Project STAR. Unfortunately, we must urge extreme caution in accepting this finding as definitive. All of the students in the sample which having no CSR experience were in the fourth grade at the time of this study, and all of the students receiving their first CSR exposure in the first grade were in the third grade when the data were collected. Thus, the differences between these groups could be due to an age-cohort difference between the third and fourth grade students. Nevertheless, the differences are significant and in favor of early exposure to small classes. By next year we will be able to determine whether the effects are reliably related to CSR experience.

Where the data presented in Table 4 present a "bottom line" view of the differences in mathematics achievement after all our potentially confounding variables are taken into account, Figure 2 traces changes in the apparent effects of CSR experience as each of the sets of contextual variables is taken into account. At the left side of Figure 2 is a cluster of five bars showing how each CSR exposure group differs from the no CSR cohort in the sample. Note that effects here are all positive and range from about one-and-a-half to nearly eight NCE points. Removing biases due to student demographic factors changes the picture very little. As the student program assignment biases are removed, however, we begin to see significant shrinkage in the apparent impact of CSR. CSR effects shrink steadily for some groups, while others do not fall into a pattern of decline until the classroom context and teacher variables are controlled. When the school effects are statistically controlled, only those groups with first grade exposure continue to display substantial positive CSR effects.

Figure 2.

Class Size Reduction (CSR) Effects on Student SAT-9 Total Mathematics NCE Achievement Scores, in Terms of Years of CSR Experience and Grade at which CSR Experience Began, after Each Block of Control Variables Was Entered in a Linear (OLS) Regression Model



Tables 5 and 6 present the same information for reading and language achievement as was presented for mathematics in Table 4. Here we see that, when the same equalization procedures are applied to achievement in reading and language, the effects of any type of exposure to CSR are very small in size and mixed in direction. The most positive benefits (though extremely small) positive effects are still concentrated on students who start CSR earlier and persist in the small classes longer.

Table 5. Average SAT-9 Reading Achievement by Class Size Reduction Experience

(NCE scores adjusted for all known implementation biases)

Starting Grade	Number of Years	Average Read. Score	Difference from No CSR	Bargraph of Test Scores
No CSR	Zero	43.15	0	
First	Two	43.79	0.64	
First	Three	43.46	0.31	
Second	One	43.77	0.62	
Second	Two	42.36	-0.79	
Third	One	41.96	-1.19	

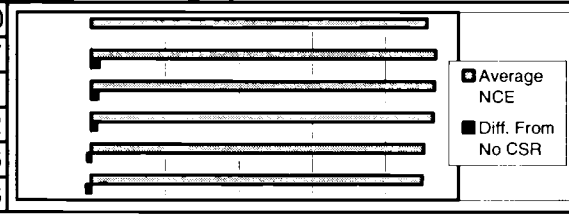
While some of the differences on this table are statistically reliable, when compared with the effects of other variables (discussed in the next section of this report) they appear truly trivial in magnitude.

The language scores reported in Table 6 present a pattern nearly identical to that for reading. Very small positive benefits for children who started CSR in first grade, with no benefit or possibly slight losses in achievement for students who begin their CSR exposure later.

Table 6. Average SAT-9 Language Achievement by Class Size Reduction Experience

(NCE scores adjusted for all known implementation biases)

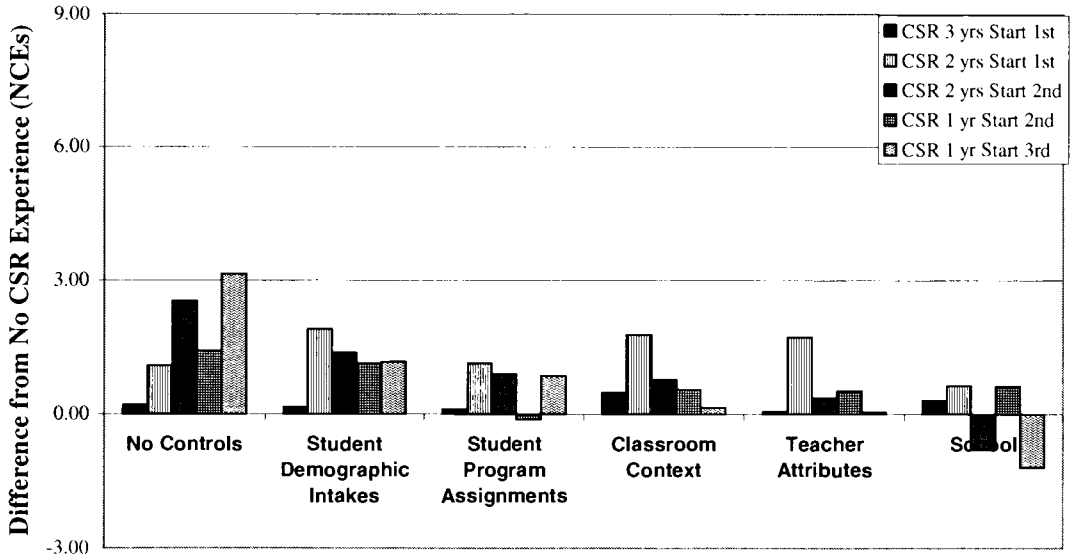
Starting Grade	Number of Years	Average Lang. Score	Difference from No CSR	Bargraph of Test Scores
No CSR	Zero	45.73	0	
First	Two	46.9	1.17	
First	Three	46.74	1.01	
Second	One	46.55	0.82	
Second	Two	45.27	-0.46	
Third	One	45.05	-0.68	



The bargraphs shown in Figure 3 trace the decline in the apparent benefits for reading achievement from reduced size class experiences as our demographic and context variables are systematically controlled.

Figure 3.

Class Size Reduction (CSR) Effects on Student SAT-9 Total Reading NCE Achievement Scores, in Terms of Years of CSR Experience and Grade at which CSR Experience Began, after Each Block of Control Variables Was Entered in a Linear (OLS) Regression Model



As is clearly visible in this figure, even before the confounding variables are subjected to statistical control, the apparent effect of CSR is much smaller for reading than for mathematics. Moreover, after all controls are applied, the negative consequences of exposure starting in the second or third grade are actually larger than the positive benefits for those beginning CSR earlier.

Taken together, these analyses of the apparent effect of CSR exposure changes when potentially confounding variables are taken into account leads to two important conclusions.

> Conclusion #3: If the California CSR initiative has created any substantial benefit to students, it only comes when students are given smaller classes beginning in their very earliest years in school.

The average across all subjects, across all of the different types of exposure to CSR beginning in the second or third grade is a negative .36 NCE points. Additionally,

> Conclusion #4: After controlling for other factors, and over all subjects, the CSR effect is never more than 5 NCE points, and averages only about 2/3 of a point. This is small compared with the negative consequences of poverty, home language and race/ethnicity (described more fully in the next section of this report).

Several factors examined in this study impact student achievement in far more dramatic ways than does class size. This does not necessarily mean that CSR is not worth the cost, however, because most of these other factors – like poverty, ethnicity and limited English fluency cannot be easily controlled, no matter how much money is spent on them.

While not discussed in detail here, the data reviewed in this report support one other important conclusion about how class size reduction might be influencing student achievement:

> Conclusion #5: There are a number of "interaction effects" indicating that exposure to reduced size classes does not benefit all students equally. These differences are not very large, and are not very consistent, however, leaving some question as to why they are appearing at all.

None of the interactions between CSR implementation patterns and student population characteristics are presented in this summary report since there is no discernible pattern to report. There is some suggestion in the data that students who had two years of reduced size class experience beginning in the first grade may have experienced more equity oriented benefits, closing somewhat the achievement "gap" between whites and other ethnic groups, but these effects are not consistent across the other CSR exposure groups raising a serious question as to whether they should be trusted at all. In an ongoing effort to reach definitive conclusions on this issue, this year's data will be subjected to more refined Hierarchical Linear Modeling analysis in the near future.

Final resolution of these questions awaits study of an additional year of data, when all of the different CSR exposure groups will have reached the fourth grade.

VII. Comparing CSR with the Effects of Other Important Variables

Throughout this report we have characterized the effects of CSR as quite small. To understand just how small they are, it helps to compare the maximum impact of any CSR experience with the achievement effects of other variables known to be influencing student achievement. Table 7 presents a summary of nineteen variables whose effects are as large or larger than the maximum class size impact shown at the bottom of the table.

Table 7. The Relative Importance of Various Factors Influencing Student Achievement on the SAT-9

	Reading Mean = 43.15 (std dev=20.12)	Mathematics Mean = 43.95 (std dev=20.96)	Language Mean = 45.73 (std dev=19.37)
Student Demographic Factors	NCE score diff	NCE score diff	NCE score diff
Loss for Afro-Americans relative to Whites	-10.66	-12.00	-9.34
Loss for Poor Students (NSLP qualified)	-8.80	-8.17	-7.80
Loss for Hispanics relative to Whites	-8.53	-8.21	-6.51
Gain for Asians relative to Whites	3.83	8.53	5.54
Loss for Other Home Language relatie to English	-6.56	-3.24	-3.23
Loss for Intradistrict mobility during the year	-3.78	-4.70	-3.97
Loss for Spanish Home Language relative to English	-5.40	-2.20	-4.03
Gain for Female Students	3.58	1.59	6.07
Classroom Assignment Factors			
Gain for being in GATE Program	21.80	23.69	21.34
Loss for being in RSP Special Educ Program	-21.62	-18.99	-17.19
Gain for Fluent Speakers relative to English Only	6.72	7.08	8.03
Loss for Overage Students (15+ months overage)	-4.79	-3.10	-4.44
Loss for being in DIS Special Educ Program	-3.17	-3.53	-4.16
Loss for Higher Grade in Combo Class	-2.68	-4.34	-2.94
Loss for Limited English relative to English Only	-4.70	-1.27	-1.32
Teacher Factors			
Loss due to not having a full credential	-1.65	-2.81	-2.33
Gain for having Hispanic rather than White teacher	1.82	1.70	1.42
Change for having a male teacher	0.03	-0.54	-1.38
Loss for each year of teacher age(a)	-0.04	-0.07	-0.05
Class Size			
Maximum Gains above Group with No CSR exposure	0.64	4.99	1.17

Note: (a) Continuous variable, the amount of change for each year.

The variables in this table are arranged in three clusters – student demographics, classroom assignment context variables and teacher factors. Within each cluster, the variables are arranged according to the overall magnitude (either positive or negative) on student performance.

In interpreting this table, it is helpful to focus first on mathematics where the maximum CSR impact was just under 5 NCE points which represents nearly a half-year of academic progress. (Remember that this is the maximum CSR impact in any subject area or type of exposure, the average mathematics gain across all CSR experience groups was only 1.57 NCE points or about 5 weeks of academic progress). The maximum 4.99 point gain is smaller than the losses incurred

due to any of four different variables: African American or Hispanic ethnicity, Poverty or assignment to a special education Resource Specialist Program. Additionally, the maximum CSR contribution to achievement is outpaced by three other variables: Asian ethnicity, enrollment in a GATE program and the advantage fluent English speakers from non-English speaking homes have over their English only peers.

Roughly comparable to the maximum CSR effect are five other variables: non-English home language, intra-district mobility, being overage for grade, being enrolled in a special education DIS program, and belonging to the upper grade in a combination grade class.

The reasons behind the alignment of achievement with this broad array of variables are quite varied, of course. The links between measured achievement and the poverty and ethnicity variables are widely documented, while their explanation remains too much a mystery. Certainly student ability has a lot to do with assignment to special education classes of various types. And it is certainly not surprising that school-to-school transiency and coming from a non-English speaking home are related to academic achievement.

Examination of the reading and language columns of Table 7 will quickly confirm that each of the factors in this table are about as strongly related to achievement in these other two domains as they are to mathematics achievement. A couple of variables are more strongly related to reading and language attainment. Girls, for example, outpace boys in reading and language by two to four times the amount of their lead in mathematics. And the losses in achievement experienced by students coming from Spanish speaking homes are about twice as great in the language related subjects. Consistency in the effects on achievement in all subjects seen across the variables listed in Table 7 contrasts sharply with the near disappearance of measurable effects on language related achievement springing from exposure to small classes.

Overall, CSR accounts for only from 0.3 to 1.3 percent of the total variance in SAT-9 test scores. The other factors reported in Table 7 explain about 40 times that amount (37.5% of math achievement, 42.0% in reading and 37.6% in language). These powerful predictors of student achievement do not take into account any contributions made by variations in student ability, prior achievement, family support, specific instructional techniques or curriculum materials.

VIII. Conclusion

Taken as a whole, this report supports eight conclusions regarding the impact of California's CSR initiative on student achievement. They include:

Conclusion #1: CSR is massive, expensive and adopted in conjunction with a complex array of other new policy initiatives aimed at improving California school performance. Evaluating the impact of this initiative is made particularly difficult by the fact that so many other important initiatives are being simultaneously pursued.

Conclusion #2: Rapid implementation of California's CSR initiative placed substantial stresses on school facilities, created an intense demand for new teachers, and encouraged a shift to Year Round school calendars to accommodate enrollment growth and reduced size classes.

Conclusion #3: School officials were faced with tough decisions regarding the sequence of CSR implementation and the allocation of opportunities to participate in reduced size classes on the part of teachers and students.

Conclusion #4: Implementation biases responsible for differences in student and teacher participation in reduced size classes were strikingly different in the first and second years of CSR implementation.

Conclusion #5: Statistical analyses revealed that biases in CSR participation are sufficiently strong that knowing the demographic, school assignment and teacher characteristics of any given student makes it possible to substantially predict whether they were in small or large classes for one or more years.

Conclusion #6: The factors associated with the biases in student participation in various CSR implementation alternatives are, themselves, much more strongly related to student achievement than is class size reduction.

Conclusion #7: Nevertheless, after controlling for all of the available biasing factors, there remains a small positive impact from CSR on student achievement as measured by the Stanford-9 achievement test. The contribution is much more powerful in mathematics that in either reading or language achievement. The CSR impact varies but is most powerful for students who were exposed during their first year in school.

Conclusion #8: Because class size reduction is so deeply entangled with student, school and teacher variables, only longitudinal analysis will make it possible to reliably disentangle the various factors influencing achievement in order to isolate the CSR contribution.

REFERENCES

Alexander, K. L., Entwisle, D. R., & Thompson, M. S. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review, 52*(5), 665-682.

Arum, R. (2000). Schools and communities: Ecological and institutional dimensions. *Annual Review of Sociology, 26*, 395-418.

Balow, I. H., & Schwager, M. (1990). *Retention in grade: A failed procedure.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics, 114*(2), 577-599.

Blatchford, P. (2000). Summary of symposium to be presented at the British Educational Research Association Annual Conference, Cardiff, 2000. Personal communication received September 28, 2000.

Blatchford, P., Goldstein, H., & Mortimore, P. (1998). Research on class size effects: A critique of methods and a way forward. *International Journal of Educational Research, 29*(8), 691-710.

Bohrnstedt, G. W. & Stecher, B. M. (Eds.). (1999). Class size reduction in California: Early evaluation findings, 1996-1998. [*CSR research consortium, year 1 evaluation report*] Palo Alto, CA: American Institutes for Research.

Brewer, D. J., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis, 21*(2), 179-192.

Burns, R. (1996). *A study of combination class achievement.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Burns, R. B., & Mason, D. A. (1998). Class formation and composition in elementary schools. *American Educational Research Journal, 35*(4), 739-772.

Clotfelter, C. T. (1998). *Public school segregation in metropolitan areas* [NBER Working Paper 6779]. Cambridge, MA: National Bureau of Economic Research.

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher, 27*(1), 5-15.

Entwisle, D. R. & Alexander, K. L. (1992). Summer setback: race, poverty, school composition and mathematics achievement. *American Sociological Review, 57*(1), 72-84.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97-109.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557-577.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (in press). The enduring effects of small classes. *Teachers College Record.*

Goldstein, H. & Blatchford, P. (1997). *Class size and educational achievement: A methodological review.* Report commissioned by UNESCO, Geneva.

Hakuta, K., Butler, Y. G., & Witt, D. (January, 2000). How long does it take English learners to attain proficiency? [*Policy Report 2000-1*]. Santa Barbara, CA: The University of California Language Minority Research Institute.

Han, L. & Hoover, H. D. (1994). *Gender differences in achievement test scores.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, Louisiana.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143-163.

Hertling, E., Leonard, C., Lumsden, L., & Smith, S. C., with contribution from L. O. Picus. (Spring, 2000). Class size. *Policy Report: Reporting on Policy Issues in K-12 Educational Management* (1), 1-20.

Hruz, T. (2000). The costs and benefits of smaller classes in Wisconsin: A further evaluation of the SAGE program. *Wisconsin Policy Research Institute Report, 13*(6), 1-44.

Hymon, S. (1997). A lesson in classroom size reduction. *School Planning and Management, 36*(7), 18-23.

Illig, D. C. (1997). *Early implementation of the class size reduction initiative.* Sacramento, CA: California Research Bureau, California State Library.

Jencks, C. & Phillips, M. (Eds.). (1998). *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics, 114*(2), 497-532.

Krueger, A. B. & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal, 111*(468), 1-28.

Mitchell, D. E., Beach, S. A., & Badarak, G. (1989). Modeling the relationship between achievement and class size: A re-analysis of the Tennessee Project STAR data. *Peabody Journal of Education, 67*(1), 34-74.

Mitchell, D. E., & Mitchell, R. E. (1999). *The impact of California's class size reduction initiative on student achievement: Detailed finding from eight school districts.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Mitchell, R. E., & Mitchell, D. E. (1999). *Student segregation and achievement tracking in year-round schools.* Paper presented at the Annual Meeting of the American Sociological Association, Chicago, Illinois.

Molnar, A., Smith, P., Zahorik, J., Ehrle, K., Halbach, A., & Kuehl, B. (December, 2000). *1999-2000 evaluation results of the Student Achievement Guarantee in Education (SAGE) Program.* Milwaukee, WI: University of Wisconsin, Milwaukee, Center for Education Research, Analysis, and Innovation.

National Conference of State Legislatures. (April, 1998). *Class Size Reduction.* Denver, CO: Author. (Available on WWW at http://204.131.235.67/programs/educ/CLASS.HTM)

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21*(2), 127-142.

———. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123-151.

Ogawa, R. T., & Stine, D. (1998). *California's class-size reduction initiative: Differences in teacher experience and qualifications across schools.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Pate-Bain, H., Boyd-Zaharias, J., Cain, V. A., Word, E., & Binkley, M. E. (1997). *STAR follow-up studies 1996-97.* Lebanon, TN: HEROS, Inc., 1997.

Rosenthal, A. S., Baker, K., & Ginsburg, A. (1983). The effect of language background on achievement level and learning among elementary school students. *Sociology of Education, 56*(4), 157-169.

Rowan, B. & Miracle, Jr., A. W. (1983). Systems of ability grouping and the stratification of achievement in elementary schools. *Sociology of Education, 56*(3), 133-144.

Shields, C. M., & Oberg, S. L. (2000). *Year-round schooling: Promises & pitfalls.* Lanham, MD: The Scarecrow Press.

Stecher, B. M., & Bohrnstedt, G. W. (Eds.). (2000). *Class size reduction in California: The 1998-99 evaluation findings.* Sacramento, CA: California Department of Education.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research, 65*(4), 319-381.

Wexler, E., Izu, J., Carlos, L., Fuller, B., Hayward, G., & Kirst, M. (1998). *California's class size reduction: Implications for equity, practice & implementation.* San Francisco, CA: PACE-WestEd.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Appendix A
Variables Analyzed in this Study

Dependent Variables. The dependent variables – reading, mathematics and language achievement – were measured using 1999 Normal Curve Equivalent (NCE) scores from the 9th Edition (Form T) of the Stanford Achievement Test (SAT-9) as mandated by the California Department of Education. In addition to assessing the specific impact of California's Class Size Reduction (CSR) initiative, this report examines the effect of student background, classroom context and teacher characteristics on individual achievement levels (i.e., Total Reading, Total Mathematics and Total Language SAT-9 scores).

Independent Variables. The central independent variable of interest in this study is, of course, class size – the number of students assigned to each teacher. We seek to determine the extent to which providing children in kindergarten through grade three with classes that have a maximum of 20 students (rather than the 28 to 32 students typical of California public schools prior to the adoption of CSR) has a positive impact on their learning. Class size is not the only influence on student learning, however. Painstaking, and often quite expensive, efforts to improve public school performance over the past several decades has taught us that student achievement is shaped by a broad range of potent demographic, social and schooling factors – factors that are often very unevenly distributed across classrooms, schools or school districts.

In the study reported here, 20 covariates with potentially powerful impacts on student academic achievement are examined. Sixteen additional variables defining classroom environmental contexts were generated by calculating classroom proportions for each factor level of seven demographic and classroom assignment variables. Taken together, these 36 variables surround and embed student achievement in five distinct contexts or levels. The five levels are depicted in Figure I. At the first level – Student Demography – five factors constitute the most fundamental and intractable academic performance influences: gender, family poverty, ethnicity, home language and time of admission to the local district.

At level 2, school organizations begin their influence on student academic opportunities by making class assignments. Five factors – grade level assignment, grade retention resulting in overage students, English language proficiency assessment, special education certification, and the level of placement (upper or lower grade) in combination grade classes – are the most obvious ***classroom assignment indicators***.

Classroom environments constitute the third context level. Classroom environments are very complex and difficult to assess precisely. They are represented in this study by several variables. Two variables of our study operate only at the classroom level – year round education track assignment and whether schools utilize combination grade classes. Additionally, this study examines fifteen calculated "concentration variables" that help to define the classroom environment by measuring the classroom proportions of:

Gender:

1. a single gender (girls),

Family income status:

2. low income status or "poverty" (children on the National School Lunch Program),

Retention in grade proxy:

3. overage-for-grade students (15+ months above a September start date for their grade),

Ethnic groups:

4. African-American (black) students
5. Hispanic students
6. Asian students
7. Other non-White students

Different home language groups:

8. Spanish home language speakers
9. Other non-English home language speakers

English language fluency groups:

10. Fluent English Proficient (FEP) students
11. Limited English Proficient (LEP) students

Special education category groups:

12. Resource Specialist Program (RSP – educationally at risk) students
13. Designated Instructional Service (DIS – blind, deaf, speech impaired, physically handicapped, etc.) students, and
14. Gifted and Talented Education (GATE) students

Intra-district transiency

15. Proportion of students in the classroom that are new to the school in the test year

Teacher characteristics comprise the fourth level of influence over student achievement. Interacting with and potentially confounding the impact of class size, we would expect to find significant influence from teacher credentials, education levels, and years of experience as well as from teacher gender, ethnicity, age and contract status.

After these variables are all controlled (using statistical procedures to remove their impact on achievement because experimental controls are not available), we would still expect unmeasured **school and district level factors** to have some influence on student achievement. At this level, we can only examine the extent to which the unmeasured influences associated with student attendance boundaries remain powerful, and to statistically remove them without having any specific explanation as to why they are affecting student test performance.

U.S. Department of Education
Office of Educational Research and Improvement
(OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Evaluating the Impact of California's Class Size Reduction Initiative on Student Achievement: Third Year Evaluation Report.	
Author(s): Douglas E. Mitchell and Ross E. Mitchell	
Corporate Source: California Educational Research Cooperative	Publication Date: March 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
↑ ☐	↑ ☐	↑ ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.		
Signature: Ross E Mitchell	Printed Name/Position/Title: Ross E. Mitchell, Research Scientist	
Organization/Address: Gallaudet Research Institute, HMB S-428 Gallaudet University 800 Florida Ave, NE Washington, DC 20002-3695	Telephone: 202-651-5576	Fax: 202-651-5746
	E-mail Address: ross.mitchell@gallaudet.edu	Date: February 4, 2002

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Educational Management
1787 Agate Street
5207 University of Oregon
Eugene, OR, 97403-5207
attn: Acquisitions